8.2 Displaying Data – Overview

Frequency Tables (Distributions)

Summarize datasets by counting the number of observations for each category, distinct value or interval.

- Can be used for categorical data and quantitative (numerical) data.

count Relative Frequency

Type of Computer	Frequency	Percent
Desktop	11	11/50 = 22%
Laptop	23	23/50 = 46%
Notebook	9	9/50 = 18%
Tablet	7	7/50 = 14%

Total 50

Number of Pets	Frequency
1	4
2	3
3	2
4	1
5	2
6	1
7	1
8	1

between 4 + 7 inclusive = 4, 5, 6, 7

Grouped Frequency Distribution

Number of Pets	Frequency
1-2	7
3-4	3
5-6	3
7-8	2

Example 1: Construct a frequency table using the data below.

38, 33, 5, 5, 47, 29, 24, 42, 3, 18,
30, 46, 25, 44, 40, 42, 39, 44, 29, 13

lower class limit upper class limit

Class	Frequency	Relative Frequency
0-9	3	3/20 = 0.15
10-19	2	2/20 = 0.1
20-29	4	4/20 = 0.2
30-39	4	4/20 = 0.2
40-49	7	7/20 = 0.35
Total:	20	20/20 = 1

100%

Class width = $Lower_2 - Lower_1$

$10 = 40 - 30$

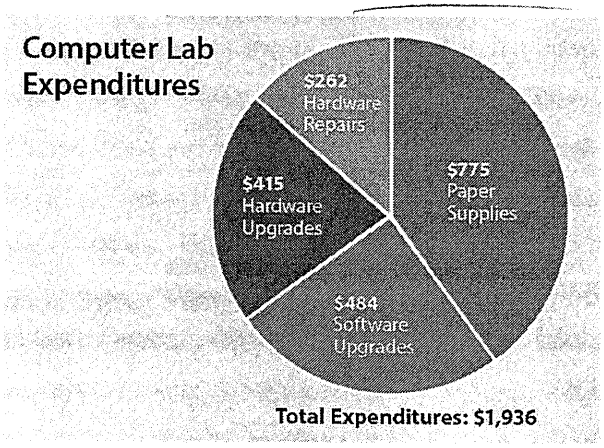
lower$_1$ lower$_2$

* All class widths must be equivalent

Graphical Displays of Data

Pie Graphs

- Compare parts to a whole.
- Slices represent the proportion of a category



Type of Data: Categorical

Advantages:

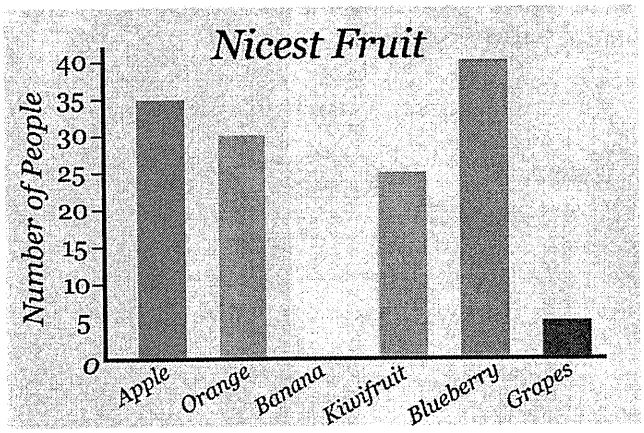
* Simple and common

Disadvantages:

* Harder to compare area than heights
* Not useful when there are lots of categories
* Easy to be misleading if visually distorted (3D, one slice is larger) or labels are not clear

Bar Graphs

- Height of the bar represents the amount of data in each category.
- Can be counts or relative frequencies.



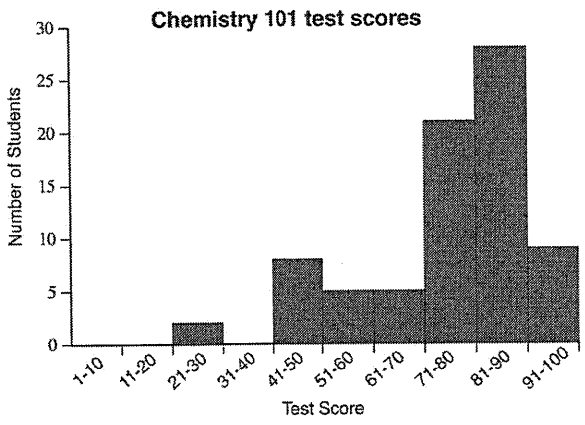
Type of Data: Categorical

Advantages:
* Simple and common and easy to read

Disadvantages:
* Misleading if:
  - Bars are not equal width
  - Inconsistent vertical scale
  - Vertical scale is truncated (not start at 0)

Histograms

- Height of the bar represents the amount of data in each ~~category~~ class.
- Can be counts or relative frequencies.



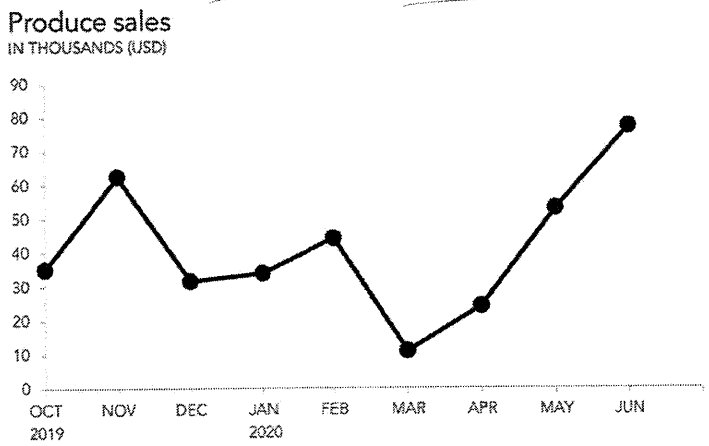
Type of Data: Quantitative

Advantages:
* Simple
* Can show lots of data very concisely
* Shows "shape" or distribution of data

Disadvantages:
* Class width impacts the plot drastically
* Misleading if:
  - Bars are not equal width
  - Inconsistent horizontal / vertical scale
  - Vertical scale is truncated (not start at 0)

Line Graphs

- Shows changes in a numerical variable over time.



Type of Data: Quantitative

Advantages:
* Shows trends over time

Disadvantages:
* Misleading if:
  - ~~Bars are not equal width~~
  - Inconsistent horizontal / vertical scale
  - Vertical scale is truncated (not start at 0)

Good

~~Misleading~~ Graphs: A clear graph should have a title, labels on the vertical and horizontal axis, and should reference the source of the data.